CLAIMS

1. A method of generating class models of semantically classifiable data of known classes, comprising the steps of:

for each known class:

extracting a plurality of sets of characteristic feature vectors from respective portions of a training set of semantically classifiable data of one of the known classes; and

combining the plurality of sets of characteristic features into a respective plurality of *N*-dimensional feature vectors specific to the known class;

wherein respective pluralities of *N*-dimensional feature vectors are thus obtained for each known class; the method further comprising:

analysing the pluralities of *N*-dimensional feature vectors for each known class to generate a set of *M* basis vectors, each being of *N*-dimensions , wherein $M << N$; and

for any particular one of the known classes:

using the set of *M* basis vectors, mapping each *N*-dimensional feature vector relating to the particular one of the known classes into a respective *M*-dimensional feature vector; and

using the *M*-dimensional feature vectors thus obtained as the basis for or as input to train a class model of the particular one of the known classes.

2. A method of identifying the semantic class of a set of semantically classifiable data, comprising the steps of:

extracting a plurality of sets of characteristic feature vectors from respective portions of the set of semantically classifiable data;

combining the plurality of sets of characteristic features into a respective plurality of *N*-dimensional feature vectors;

mapping each *N*-dimensional feature vector to a respective *M*-dimensional feature vector, using a set of *M* basis vectors previously stored, wherein $M << N$;

comparing the M-dimensional feature vectors with stored class models respectively corresponding to previously identified semantic classes of data; and

identifying as the semantic class that class which corresponds to the class model which most matched the *M*-dimensional feature vectors.

3. A method according to any of the preceding claims, wherein the set of semantically classifiable data is audio data.

4. A method according to claims 1 or 2, wherein the set of semantically classifiable data is visual data.

5. A method according to claims 1 or 2, wherein the set of semantically classifiable data contains audio and visual data.

6. A method according to any of the preceding claims, wherein the analysing step uses Principal Component Analysis (PCA).

7. A method according to any of claims 1 to 5, wherein the analysing step uses Kernel Discriminant Analysis (KDA).

8. A method according to any of the preceding claims, wherein the combining step further comprises concatenating the respectively extracted characteristic features into the respective *N*-dimensional feature vectors.

9. A system for generating class models of semantically classifiable data of known classes, comprising:

feature extraction means for extracting a plurality of sets of characteristic feature vectors from respective portions of a training set of semantically classifiable data of one of the known classes; and

feature combining means for combining the plurality of sets of characteristic features into a respective plurality of *N*-dimensional feature vectors specific to the known class;

the feature extraction means and the feature combining means being repeatably operable for each known class, wherein respective pluralities of *N*-dimensional feature vectors are thus obtained for each known class;

the system further comprising:

processing means arranged in operation to:

analyse the pluralities of *N*-dimensional feature vectors for each known class to generate a set of *M* basis vectors, each being of *N*-dimensions , wherein $M << N$; and

for any particular one of the known classes:

use the set of *M* basis vectors, map each *N*-dimensional feature vector relating to the particular one of the known classes into a respective *M*-dimensional feature vector; and

use the *M*-dimensional feature vectors thus obtained as the basis for or as input to train a class model of the particular one of the known classes

10. A system for identifying the semantic class of a set of semantically classifiable data, comprising:

feature extraction means for extracting a plurality of sets of characteristic feature vectors from respective portions of the set of semantically classifiable data;

feature combining means for combining the plurality of sets of characteristic features into a respective plurality of *N*-dimensional feature vectors;

storage means for storing class models respectively corresponding to previously identified semantic classes of data; and

processing means for:

mapping each *N*-dimensional feature vector to a respective *M*-dimensional feature vector, using a set of *M* basis vectors previously generated by the third aspect of the invention, wherein $M << N$;

comparing the M-dimensional feature vectors with the stored class models; and

identifying as the semantic class that class which corresponds to the class model which most matched the *M*-dimensional feature vectors.